University of Illinois Graduate School of Library Science

OCCASIONAL PAPERS

No. 101 February 1972

EVALUATION OF ON-LINE SEARCHING IN MEDLARS (AIM-TWX) BY BIOMEDICAL PRACTITIONERS*

by

F. Wilfrid Lancaster

TABLE OF CONTENTS

INTRODUCTION

The purpose of the investigation was to determine how effectively biomedical practitioners, with a minimum of introduction to the system, can conduct on-line searches to satisfy their own information needs. The searches were conducted in the Abridged Index Medicus data base as implemented on the on-line ELHILL system (AIM-TWX). ELHILL is the ORBIT on-line retrieval system of the System Development Corporation as modified for National Library of Medicine use. AIM-TWX was a particular experiment whereby ELHILL was used to make the data base of Abridged Index Medicus available by teletypewriter exchange to medical centers in various parts of the country. The searches used in the study were conducted at four MEDLARS centers having the on-line search facilities available, in the period November 1970 - February 1971. The size of the AIM-TWX data base was approximately 100,000 citations in this period.

* The investigation reported in this paper was conducted by the author for the National Library of Medicine and is printed with the permission of the National Library of Medicine.

PROCEDURES

MEDLARS search analysts at the participating centers were asked to identify health professionals coming to the center with requests for information whose needs might be satisfied, at least in part, by the AIM-TWX data base, and who indicated willingness to cooperate in the study. Each of these professionals was asked to complete a special search request form (see Fig. 1).

The search analyst introduced the practitioner to AIM-TWX by presenting for his examination a very brief printed description of how to use the system. This description was an outline, with text and examples, designed to give the requester just enough information to allow him to sit down at the terminal and conduct relatively simple searches. No attempt was made to describe all the ramifications and sophistications of ELHILL. The description was prepared at the National Library of Medicine (NLM) and was standard for all users. The search analyst was allowed to answer any questions the user might have after reading the instruction summary, but he was not to volunteer any additional clarification.

Having read the summary the user was left to conduct his own search at the terminal. Logging-in to the system, a simple technical procedure, was done for the user. The user had available to him: Medical Subject Headings (MeSH), a list of subheadings and their abbreviations, a list of check-tags, the MeSH vocabulary in hierarchical form (Tree Structures), and the full AIM-TWX user guide (for reference purposes only).

The search analyst was instructed to remain in the general vicinity of the terminal. She could answer any technical questions that arose during the search and could assist the user with "technical problems" (i.e., problems relating to the equipment and the on-line protocols, but not problems relating to MEDLARS indexing, vocabulary or search strategies). The user was told that he was participating in an experiment and that the experimental constraints did not permit his receiving help with the intellectual aspects of the system or the search. In other words, a situation was being simulated in which the user would be at a remote terminal with no trained searcher available to give advice and guidance.

A user's search was considered to be completed when either: (1) he decided he had found enough references, or (2) he gave up the search after trying various approaches. The searcher then logged-out of the system for the user.

When the search was completed the user was asked to examine each citation that had been printed in his on-line search and to mark each with a code to indicate its relevance to his information need, on a scale which included the designations "relevant," "peripherally relevant," and "irrelevant." Where necessary, copies of the actual articles were made available to allow unequivocal assessments.

A copy of the entire search dialog was made (the user retained the original), and on this copy the analyst recorded the exact time spent by the user at the terminal (available from log-in and log-out times) and the user's subjective assessment of the value of the search to him, on the scale "major value," "considerable value," "minor value," or "no value."

The search analyst, after examining the user's results, then conducted

Lib Sci

1W

Name of requester:

Title:

Organization:

Department:

Telephone:

M.D., Ph.D., RN, other:

Major area of responsibility: Basic research

Clinical research

Teaching

Clinical practice

Other (please specify)

Number of times requester used AIM-TWX:

Subject of present search (Record below, in your own words, the subject matter for which the search is to be conducted. Be as specific as possible):

Purpose of search: Clinical problem

On-going research

Writing a research paper

To assist in: Writing a review article or book chapter

Preparing lecture or other teaching function

Other (please specify:) ____________________

____

Type of search:

Are you looking for: a. A few relevant papers

b. All possible relevant citations

Fig. 1. AIM-TWX Search Request Form

her own search on the topic at the AIM-TWX terminal. The object was to try alternative approaches to retrieval in an effort to find additional relevant citations that the user might have missed or to find citations to articles that might be more highly relevant than those found by the user. Any new citations thus discovered were submitted to the user for relevance assessment on the same scale as that previously used.

Analysis of the user's search was conducted on the basis of his request form, the dialog of his interaction with the system, his relevance assessments, his search time, his judgement on the value of the search, and the results of the parallel search conducted by the analyst.

RESULTS

In all, 48 test searches were completed, although full data were not gathered for every one of these. The majority came from Ohio State University with a few from UCLA, NLM and Alabama. The numerical results are presented in Table 1. For each search the following data are presented:

1. The number of times the requester had previously used AIM-TWX (although some of these data are suspect since apparently some users checked the "previous use" box if AIM-TWX had been used on their behalf by a MEDLARS analyst).
2. Whether he wanted to find all relevant citations or only a few of them (high or low recall).
3. The precision ratio of his search, i.e., the number of unique relevant citations among the total number of unique citations printed.
4. The unit cost per relevant citation retrieved which is found by dividing the total number of unique relevant citations found into the total time spent at the terminal. Where no relevant citations were found by the user (as in search #1) it is not possible to express a unit cost. In these cases the total search time is quoted.
5. The recall estimate for the user's search which is derived from the equation:

$$\frac{\text{Number of unique relevant references found by user}}{\begin{array}{c}\text{Number of unique relevant references found by user and number of}\\ \text{additional unique relevant references found by analyst}\end{array}}$$

In some cases no parallel search was conducted through error or lack of time. These searches are indicated by a dash in the recall column. In other cases the search analyst could think of no further approaches so did not undertake a parallel search. For these searches a single asterisk appears in the recall column. In actual fact, one might equally well write in 100% recall for these. A cross (+) mark is used to indicate recall figures that are definitely suspect. These are searches in which the analyst printed only a selection of the citations retrieved by a strategy. Other citations, matching the strategy but not printed, are almost certainly relevant. Because they were not printed, however, and thus not assessed for relevance, these citations are not included in the recall estimates. In other words, the recall estimates quoted in these searches are almost certainly inflated.

The final column of Table 1 records the user's subjective assessment of the value to him of his own search, on the scale: major value, considerable value, minor value or no value. For some searches, however, the responsible analyst

TABLE 1

Search No.	No. of Previous Uses	Recall Requirement	Precision	Unit Cost	Recall Estimate	Value
1	3	all	0/5=0	----	0/0 (?)	none
2	0	few	4/10=40.0	2.5 min. total	----	----
3	1	few	6/7=85.7	1.8	----	----
4	2	few	8/10=80.0	1.2	----	----
5	0	few	7/10=70.0	4.3	7/7=100%	major
6	1	all	15/16=93.7	3.2	15/23=65.2	considerable
7	1	few	0/1=0	----	0/0 (?)	none
8	1	few	0/11=0	---	0/2=0	none
9	4	few	0/0	---	*	none
10	2	few	6/6=100	2.0	6/8=75.0	considerable
11	0	few	11/11=100	2.3	11/24=45.8	considerable
12	2	all	3/3=100	4.3	3/3=100	----
13	0	all	12/12=100	2.0	12/12=100	major
14	0	few	10/13=76.9	2.0	10/19=52.6	considerable
15	0	all	6/8=75.0	4.3	6/9=66.7	minor
16	0	all	1/5=20.0	26.0	1/9=11.1+	minor
17	1	all	0/5=0	----	0/11=0	none
18	3	few	1/1=100	4.0	1/3=33.3	major
19	0	few	0/0	----	0/6 =0	none
20	1	few	0/7=0	----	0/1=0	none
21	1	all	9/19=47.4	5.5	9/9=100	considerable
22	4	few	29/29=100	1.0	----	major
23	0	all	4/6=66.7	7.0	----	minor
24	0	all	27/27=100	1.4	27/45=60.0	considerable
25	0	all	3/7=42.8	10.0	3/12=25.0	minor
26	2	all	2/2=100	12.0	2/3=66.7	considerable
27	5	few	7/7=100	1.4	*	major
28	0	few	12/23=52.2	2.2	12/13=92.3+	major
29	1	few	5/8=62.5	3.8	5/25=20.0	considerable
30	20	few	5/10=50.0	3.4	*	major
31	7	all	16/18=88.9	1.2	*	major
32	0	few	21/30=70.0	1.0	21/33=63.6+	----
33	0	few	6/7=85.7	3.0	6/10=60.0+	considerable
34	1	all	6/11=54.4	5.0	6/13=46.1	major
35	0	all	8/8=100#	3.0	8/8=100#	----
36	2	all	5/7=71.4	6.0	5/13=38.4	----
37	0	all	17/19=89.5	3.4	17/35=48.5	major
38	0	all	6/13=46.1	8.0	6/8=75.0	major
39	1	all	0/2=0	----	0/25=0	none
40	0	all	23/24=95.8	5.2	23/45=51.1	considerable
41	0	all	29/70=41.4	3.8	29/40=72.5	minor
42	0	all	4/14=28.6	12.5	4/13=30.8	considerable
43	2	all	10/11=90.9	1.5	10/11=90.9	considerable
44	0	all	13/25=52.0	6.1	13/22=59.0	----
45	0	all	10/18=55.5	6.0	10/16=62.5	considerable
46	0	few	5/9=55.5	----	5/12=41.6	----
47	0	all	39/54=72.2	2.4	39/40=97.5	considerable
48	1	few	115/143=80.4	0.25	115/226=50.9	major

* Search analyst could think of no alternative approaches
\+ Doubtful recall figure
\# Search on author name

failed to obtain the user's assessment of value. In these searches a dash appears in the appropriate column.

The average precision figure for the group of searches was 63.1%. This is based on 45 of the 48 searches. Two searches with zero retrieval were omitted from the precision calculation, as was search #35 which was an author search and therefore could not reasonably be expected to score less than 100%. In other words, on the average, over 60% of all the citations retrieved were judged either relevant or peripherally relevant and less than 40% of those printed were judged to be not relevant. I consider it extremely encouraging that a group of biomedical specialists with a minimum of exposure to the system (one-half the group had never used AIM-TWX before and most of the remainder had used it only on one or two occasions) should be able to conduct searches to satisfy their own needs, using a controlled vocabulary, and achieve a precision in excess of 60%. In terms of precision there were few really bad searches. Searches #1, #7, #8, #17, #20, and #39 achieved zero precision. That is, they retrieved some citations (the worst, #8, retrieved 11) but none were relevant. However, in the case of two of these searches the MEDLARS analyst was unable to find any relevant citations so that for these there is a strong possibility that no relevant citations exist in the AIM-TWX data base. Nine of the searches achieved 100% precision. It is worth remembering that in the full evaluation of MEDLARS, reported in 1968, the average precision for 300 off-line searches, conducted by trained analysts, was only 50%.[1]

Another interesting figure is the unit cost (in time) to the user per relevant citation retrieved. This figure is obtained by dividing the total time at the terminal by the number of relevant citations retrieved and is a valid measure of the cost to the user in finding relevant references. The unit cost is available for 39 of the searches. For 8 searches no relevant citations were retrieved and it is not possible to express a unit cost for these (unless we regard them as infinite); one search terminal time was not recorded so this too is omitted. For the 39 searches the unit costs range from a high of 26.0 minutes (search #16 took 26 minutes and retrieved only one relevant citation) to a remarkable low of 0.25 minutes (search #48 discovered 115 relevant citations in 29 minutes). The average unit cost, over the 39 searches, is 4.5 minutes per relevant citation, and the median unit cost is 3.4 minutes. For 30% of the searches the unit cost per relevant citation was 2 minutes or less. Considering that most searches were reasonably complex, requiring coordinations between two or more aspects, these search times appear to be good and well within tolerable limits.

The most reliable average recall figure is based on a group of 36 searches. This group excludes 5 searches for which, through error, no parallel analyst searches were conducted. It also excludes 2 searches for which no relevant references were retrieved by either search, as well as 4 searches in which I feel the recall estimate is suspect (for reasons mentioned above) and one search conducted on the name of an author only (where it would be very dificult to get less than 100%). The group includes, however, 4 searches in which the search analyst felt unable to improve on the user's search and which were all scored 100% recall for the user. For the group of 36 searches the average recall was 57.6%. This estimate is almost identical to the average recall (57.7%) achieved over 300 searches in the full MEDLARS study. Recall estimates range from 100% for 7 searches to zero recall for 5 searches. The worst recall result occurred in search #29 where the user was unable to find any relevant citations but the analyst found 25.

An overall average recall estimate of 57% appears entirely satisfactory for these searches particularly when we consider that almost half the requesters (22 of 48) indicated that they were looking for a few citations only. To these requesters recall is unimportant. In fact, if we take the group of 22 searches for which high recall was required, and for which there appears to be reasonably good recall estimates, the average recall is somewhat higher, 61.7%, than for the entire group of 36 searches.

Perhaps the best indication of the success or otherwise of a search is the user's own subjective assessment of its value to him. Value judgments were obtained for 39 or the 48 searches, with the following results:

TABLE 2

Scale Values	Number	Percent
MAJOR VALUE	12	30.8
CONSIDERABLE VALUE	14	35.9
MINOR VALUE	5	12.8
NO VALUE	8	20.5

Again, I interpret this as being a satisfactory result, especially when we consider that in the case of two of the "no value" searches the result appears to reflect the absence of relevant citations in the AIM-TWX data base (both analyst and user were unable to find any relevant items) rather than defects in searching strategies. That 67% of the users judged their searches of considerable or major value is indeed encouraging for the future of on-line searching by biomedical practitioners.

It is interesting to consider the possible effects of the learning process on the results achieved at the terminal. Dividing the searches into three groups, by number of previous uses of AIM-TWX, one arrives at the following results:

TABLE 3

	Precision	Recall	Unit Cost
System not used before (24 searches)	62.4%	58.7%	5.4 min.
Used once or twice before (17 searches)	56.8%	46.6%	4.1 min.
Used more than twice before (7 searches, 5 of which were judged of major value by the user)	62.7%	86.6%	2.2 min.

No particularly clear picture emerges from these figures, except that increased experience appears to reduce the unit cost per relevant citation retrieved. This is to be expected: as experience at the terminal increases, less time is wasted in entering invalid terms or using incorrect constructions. From Table 3 it can be seen that the most experienced group of users achieved the best results in terms of recall, precision and unit cost. However, the group of users who had searched AIM-TWX on one or two occasions before the test search actually performed less well, in terms of recall and precision, than the group who had not previously used the system (although as previously mentioned, some "previous uses" were actually previous uses by an analyst for the user). Unfortunately these samples are too small to indicate anything very clearly and there are too many other variables affecting the results, besides the variable of learning. One of these variables is the complexity of the request.

It is more revealing to compare the characteristics of a group of "worst" searches with the characteristics of a group of "best" searches. When we examine the searches with high recall and precision results, we see that most have at least the following features:

1. They involve relatively simple relationships.
2. The terms of the request statement translate fairly directly into MeSH headings. Some examples include the following:

   #12 Cryosurgery of rectal carcinoma
   (Searched on CRYOGENIC SURGERY and RECTAL NEOPLASMS)

   #13 Histocytochemistry and electrophoresis of creatine kinase
   (Searched on CREATINE KINASE and HISTOCYTOCHEMISTRY or ELECTROPHORESIS)

   #27 Reviews on childhood schizophrenia
   (Searched on SCHIZOPHRENIA, CHILDHOOD and REVIEW)

The poor searches, on the other hand, may be divided into two categories:

(a) Requests for which nothing of more than peripheral relevance is found in the data base but a lot of time is wasted in establishing this. Examples are search #8 and search #20, respectively, on the skin punch biopsy technique in plastic surgery and the selection of patients for rhinoplasty. In these cases poor results are due to the fact that little relevant literature exists, at least in this data base, and the searcher cannot be blamed for this.

(b) Requests that require more sophisticated search techniques, either because they involve more complex conceptual relationships or because it is by no means obvious, to the inexperienced searcher, what the appropriate MeSH terms should be. A good example is search #19 on "hyperalimentation" which was searched on this term (not a MeSH term) but needed other combinations (e.g., PARENTERAL FEEDING and WATER-ELECTROLYTE BALANCE) for a successful search. The inexperienced user could hardly be expected to know this and here the system is obviously at fault. An on-line system used by health professionals needs a full entry vocabulary. The AIM-TWX system has no built-in entry vocabulary and MeSH itself is grossly lacking in specific entry terms. Terms such as "hyperalimentation" that describe concepts upon which literature exists in the system should appear in an on-line entry vocabulary, with appropriate mappings to MeSH terms or combinations. (It is noteworthy that a rather complete entry vocabulary is an important feature of on-line experiments at the British MEDLARS center.)

An example of a more complex search, in which the user had little success, is #39, relating to solutions used in automatic clinical blood cell counters. The user tried only BLOOD CELL COUNT and SOLUTIONS, whereas most of the relevant citations were retrieved by the MEDLARS search analyst on ERYTHROCYTE COUNT/INSTRUMENTATION or BLOOD CELL COUNT/INSTRUMENTATION.

Having examined 48 searches in some detail it is possible to make some generalizations on the entire group of test searches as conducted by biomedical practitioners who are not information specialists:

1. Searches tend to be quite effective and efficient where the conceptual relationships are not highly complex and where MeSH terms match request terms fairly closely.
2. Lack of an entry vocabulary seriously hampers the user in cases where his request terms have no exact equivalent in MeSH.
3. Users are most successful in relatively simple approaches. When they try more sophisticated techniques they frequently go astray (e.g., asking for an explosion on a main heading/subheading combination).
4. A major cause of low recall is the user's failure to recognize all possible approaches to retrieval. For example, in search #46, relating to the post-gastrectomy "dumping" syndrome, the user tried DUMPING SYNDROME/THERAPY but overlooked DUMPING SYNDROME/SURGERY.
5. The interactive capabilities of the system are comparatively little used. Frequently the searcher will stick to his original search strategy and will not be led to alternative approaches.
6. Very few users choose the "print full" option. Those who do, however, generally find it useful to do the following:

   (a) successively narrow the scope of a search (e.g., TRANSPLANTATION and not TISSUE PRESERVATION and not DIALYSIS and not HISTOCOMPATIBILITY),
   (b) expand the scope of a search by discovering new terms, and
   (c) expand the scope of a search by discovering new facets of interest.

Before leaving these results we should consider the composition of the group of users represented in the evaluation. About a third of the group are M.D.'s, including associate and assistant professors, residents, interns and postdoctoral fellows; ten users are medical students (third or fourth year); and most of the remainder are research associates or assistants (e.g., in microbiology), although one is a physical therapist and another is president of a company manufacturing medical equipment.

By "major area of responsibility" the users characterized themselves as involved in the following:

Clinical practice	14
Basic research	11
Clinical research	10
Teaching	9
Drug information service	2
Product development	1

The "purpose of search" was indicated as follows:

Clinical problem	18
Ongoing research	10
Writing research paper	7
Writing review article or book chapter	6
Preparing lecture or teaching function	6
Writing book or thesis	2
Preparing grant application	2

These categories are not mutually exclusive because some users checked more than one category. For example, a user might indicate two major areas of responsibility, such as teaching and basic research. Likewise, the purpose of a search might be to support ongoing research and also to help in writing a research paper.

OBSERVATIONS AND CONCLUSIONS

In this section I present my personal interpretation of the results of this experiment and discuss the implications of the results for on-line searching in the MEDLARS data base by biomedical specialists. Limitations of the present ELHILL system will be discussed as well as possible longer-term approaches to the problems of on-line retrieval.

First it will be worthwhile to consider the disadvantages of off-line, batch-processing for information retrieval purposes. The major limitations are that:

1. There is very little possibility for browsing.
2. A search strategy cannot be developed heuristically. The searcher has essentially one chance to conduct a successful search and must therefore think in advance of all reasonable approaches to retrieval.
3. The search must be delegated to an information specialist. The patron of the information service is not able to conduct his own searches. Unfortunately delegation causes problems. Users sometimes have difficulty in describing what they are seeking and search analysts may misinterpret a user's requirements. These problems exist in all delegated search systems and were well documented in the evaluation I conducted on the full MEDLARS system.[1]
4. There is a time delay. In a batch-processing system there is an inevitable delay in obtaining search results. There is certainly no opportunity for "real time" response.

The on-line search system has none of these disadvantages. Even for delegated searches conducted by trained analysts, the on-line mode of implementation has the advantages of rapid response and the capability for interactive, browsing, heuristic searches. Ultimately, however, on-line systems should be capable of being used in a non-delegated search mode. That is, the practitioners in a field should be able to undertake productively their own literature searches without the interposition of an information specialist. The problems of misinterpretation and miscommunication are thereby avoided. It is this aspect, the use of an on-line search facility by practitioners of medicine, that I am concerned with here.

On the whole I believe the results to be surprisingly good. Although a few users went badly astray, the majority were able to conduct productive searches. The precision achieved in most cases was high and the cost in time appears to be well within tolerable limits. It is not to be expected that the requester himself will perform as well as would a trained analyst. He probably would not use Index Medicus as effectively either. He cannot master the complete vocabulary, indexing policies and the niceties of search strategy in a matter of minutes. Nevertheless, acceptable results were obtained in most cases by a simple and straightforward approach. It is noteworthy that many of these users were seeking only a few relevant references. They did not require high recall. Frequently they were satisfied with the results of their own efforts and did not really need the additional analyst search, which was, however, conducted anyway for the purpose of the study. For the user who would like everything in the data base, the search analyst will usually be able to find some additional references because she is able to think of alternative approaches. It is also noteworthy, however, that in several cases the searcher was not able to find any additional material and in some instances did not even try because she felt there was little possibility of improving the user's search.

Although it would be desirable to have more evidence on this point, by means of additional searches conducted with the same controls (I asked for a minimum of 100 for evaluation purposes but received only 48), I am convinced from these results that many a biomedical practitioner of the type encountered in this experiment can exploit AIM-TWX profitably with the minimum of introduction to the system and without the necessity for having a trained MEDLARS analyst at his elbow. In other words, it would be perfectly practical to provide the AIM-TWX service in medical centers (where there is an indicated demand) remote from MEDLARS search facilities. However, it would always be desirable to have one or more people at this center (e.g., the librarian) more fully familiar with the system and its capabilities. In particular someone should be knowledgeable in the purely technical aspects of logging-in, logging-out and recognizing and dealing with technical problems as they arise. If the terminal were serviced by a computer facility such technical matters could be handled adequately by the staff of this center. It would also be useful if a MEDLARS analyst could be available, perhaps via telephone, to help with real problems.

What is still lacking is a brief, clear, well illustrated (with examples) booklet describing in simple terms how to use the system. This publication should not attempt to present everything. Not all command options are of equal importance and some will rarely be used. It is more important to present the essentials of the system and also to illustrate how the practitioner can use MeSH, the hierarchical Tree Structures (which is potentially his most valuable tool) and any other aids that are deemed desirable (e.g., lists of check tags). Given such a publication I have little doubt that AIM-TWX could be used profitably by many biomedical professionals. Undoubtedly, however, a live demonstration has potentially more value than any guide.

AIM-TWX appears to meet a definite need. The great majority of the searches conducted as part of this experiment could not have been conducted in Index Medicus. They involve conceptual relationships that would be virtually impossible to handle without some facility for term coordination. Moreover, in many cases the searches involved terms that would not necessarily be print terms.

AIM-TWX, then is not made redundant by Index Medicus (IM) or Abridged Index Medicus (AIM). Moreover, it appears that the system is frequently used for searches that would probably not be conducted in MEDLARS by the usual off-line route--the type of search in which the requester seeks a few references and needs them right away. The on-line system, then, serves a function that is not fulfilled by either the full MEDLARS demand search service or the various printed products of MEDLARS. Further, search analysts at the participating centers have become aware that the AIM-TWX facility is attracting users who have not previously requested MEDLARS searches and, in fact, have not been regular users of the medical library in the past.

ELHILL is a good system of its type. It includes various features (e.g., the explosion capability) not always found in on-line search systems. It is relatively easy to use. However, it can certainly be improved. I have examined in detail the 48 test searches and compared them with the parallel searches conducted by the MEDLARS analyst. I have attempted to identify major problems and to determine what might be done to assist the on-line user in preparing better search strategies. I present my observations and recommendations below, dealing first with ELHILL as such and secondly with on-line MEDLARS searching in general.

POTENTIAL IMPROVEMENTS TO ELHILL

The principal limitation of ELHILL is that it requires virtual perfection in the entering of search terms and commands. Although it is relatively easy to make corrections, the system makes little attempt to compensate for human error. Unfortunately it is very easy to make simple mistakes, particularly if one is not used to typing or other keyboarding activities. Some common types of error are illustrated in the following real examples:

LYSERGIC ACID DIETHYLAMIDE,	Superfluous comma
HEMORRHAGE POSTPARTUM	Comma missing
LONG ACTING THYROID STIMULATOR	Hyphen missing
RECTAL NEOPLASM	Final 's' missing
DEXTRO AMPHETIAMINE SPLENIC ARETERY TRANSPLANTATHON	Spelling error
MYCOTIC ANEURYSM	Transposition. Should be ANEURYSM, MYCOTIC
TRANSPLANTATION/SKIN	Main heading/subheading transposed

In none of these cases, except possibly the last, should the search terms be rejected. The system should be programmed to compensate for and accept simple errors such as the omission of a hyphen or a comma, or the insertion of a superfluous punctuation mark. It should also be programmed on the principle of "minimum character string recognition." The string DEXTRO A ... is sufficent to identify the heading of DEXTRO AMPHETAMINE and to distinguish it from every other descriptor presently in the vocabulary. DEXTRO A should be acceptable by the system even though a typing error occurs later in the string.

Simple errors of these kinds frequently occur and should not be cause for complete rejection of a descriptor. The inexperienced user has enough to tackle in grasping the searching methods and commands as well as familiarizing himself with the vocabulary. He should not be bothered with the need for perfection in spelling and punctuation. Term rejection for trivial reasons is irritating and unnecessary and reduces user tolerance to the system.

A somewhat annoying feature of AIM-TWX is the duplication of relevant citations. If a citation matches several term combinations tried by the searcher, it will be printed several times in the course of a search, unless the searcher successively negates all previous combinations used. It should be possible to avoid this step in the search programs themselves by storing a record of all citations printed in a particular search and not printing these again during the course of the same search unless this feature is overridden. In this case the tallies displayed would reflect only new citations matching a strategy rather than the total of citations matching.

An on-line system, particularly if it is used by people who are not trained analysts, needs a large well-prepared entry vocabulary. Indeed, the entry vocabulary is a very important part of any retrieval system.[1] At present the AIM-TWX system has no entry vocabulary. It will accept only recognized MeSH headings and subheadings. Even the see references in MeSH are omitted. It should be possible for a user to search on any entry term occurring in MeSH or in the Integrated Authority File and have the transformation to the correct descriptors conducted automatically. For example, a user should be allowed to enter such terms as HOMOGRAFTS or GRAVE'S DISEASE and have these terms converted respectively to TRANSPLANTATION, HOMOLOGOUS and GOITER, EXOPHTHALMIC. Word transposition should also be recognized and corrected via the entry vocabulary or by general rule (e.g., the MYCOTIC ANEURYSM versus ANEURYSM, MYCOTIC situation). Main heading/subheading transposition may be handled in the same way, although this is more difficult and is less critical. In the on-line implementation of MEDLARS in England, facilities for the recognition of entry terms are built in, and this is as it should be.

An on-line system does not necessarily eliminate completely the need for desk-top tools. This is particularly true if the system is implemented by a typewriter terminal without cathode ray tube display. There are certain things that are better done by printed tool than by on-line printout. For example, to conduct a generic search in a category we need to view the appropriate section of the term hierarchy (to confirm it does indeed cover the aspect we are looking for) and we need to obtain the number of the term under which an explosion is to be conducted. It is easier and more useful to examine the hierarchy and read off the necessary numbers in the printed Tree Structures than to go through the process of entering successively the commands TREE (which in any case gives only a small segment of the hierarchy), MESHNO and then eventually EXPLODE. In search #37 the user wasted considerable time trying to get MeSH numbers for terms (and making several typing errors in the process) and probably becoming very frustrated. He could have done the whole thing much more easily in the printed tool (using MeSH as an index). Moreover, use of the printed tool avoids wasting time on useless explosions. At least one user, for example, exploded terms which have no subclasses in the system. He thereby merely deluded himself and wasted his own time since he had searched the terms unexploded earlier. The on-line system should not do things that are better handled by a printed tool and AIM-TWX should therefore de-emphasize these aspects. Instead, more time should be spent in developing improved procedures for functions that cannot be handled by the printed book approach.

This is not a condemnation of the ELHILL system which, as I said earlier, is an excellent example of its type. I am merely pointing to ways in which the system should be improved to make it easier to use and to reduce the burden on the inexperienced searcher at the terminal.

ON-LINE SEARCHING OF MEDLARS IN GENERAL

In a few paragraphs I will try to apply my analysis of the test searches to the subject of on-line MEDLARS searching, by health professionals, in general. I will endeavor to indicate what a future on-line MEDLARS (not necessarily AIM-TWX) should be like if it is to be of maximum utility to the biomedical practitioner (using it directly).

First we have to recognize that, historically, MEDLARS is a system designed to be used primarily by trained analysts. It was not designed with the non-delegated, user search in mind. True, the publications produced are intended for general consumption, but the retrospective search aspect is very much oriented toward use by information specialists. Until comparatively recently no entry vocabulary was widely available. In fact, there is still no extensive entry vocabulary for general public use. Index Medicus is lacking an adequate cross-reference structure and Medical Subject Headings itself contains only a minimum of entry terms. Moreover, MeSH does not present the hierarchical structure (by "broader term" and "narrower term" references) that is displayed in the conventional thesaurus. On the other hand, the Tree Structures, which provide the full hierarchy in the best possible form, are not widely available to the general public (or even to medical librarians). Fully effective use of MEDLARS requires a knowledge of a large controlled vocabulary (which changes with time), of indexing policies and protocols, and of techniques for combining terms in various ways to produce searching strategies. This knowledge is not gained overnight: witness the length of the present analyst training program.

Since MEDLARS has been designed for use by information specialists, and these specialists require extensive training, it follows that the user not so indoctrinated will require much greater help at the on-line terminal than will the trained analyst. And he should receive much greater help than that provided by the present AIM-TWX system. This does not mean that the untrained person cannot use the MEDLARS data base productively--the present study has indicated that he can. It means that we should strive to produce improved systems that are more user-oriented and that will help the user to attain higher levels of success. Some ways in which these goals might be achieved are outlined below.

The difference between an analyst search and a non-analyst search can be exemplified by considering search #37. The topic was myoglobinuria occurring in hypokalemia induced paralysis in humans. The practitioner used a very simple search approach:

MYOGLOBINURIA		
HYPOKALEMIA	and	PARALYSIS

This simple approach was relatively effective. Nineteen citations were retrieved and seventeen of these (precision ratio of 89.4%) were judged relevant. However, the search analyst was able to find another eighteen relevant citations using more sophisticated searching strategies as follows:

	1		2
1.	MYOGLOBINURIA		HYPOKALEMIA
	PROTEINURIA		POTASSIUM/METABOLISM
	PROTEINS/URINE	and	POTASSIUM/PHYSIOLOGY

2. 2 and PARALYSIS (explosion) and HUMAN

3. PARALYSIS (explosion) and 1 and HUMAN

4. (1 or 2) and HUMAN and ELECTROMYOGRAPHY
   MUSCLES/PATHOLOGY
   MYOTONIA

In this case the user indicated a need for all relevant citations so the analyst search was clearly superior. The question that arises is "What could be done to help the biomedical practitioner achieve search results more comparable to those achieved by the trained analyst?" In actual fact, with a well constructed network of cross-references and an adequate entry vocabulary, the problem may not be as great as it seems at first sight. In this search, for example, many of the relationships used by the search analyst are displayed explicitly in MeSH. Using the terms he began with, the user might reasonably have been led by MeSH itself to a consideration at least of several of the additional terms used by the analyst. MYOGLOBINURIA is shown to be related to PROTEINURIA, PROTEINURIA is referred to from URINE, and HYPOKALEMIA is referred to (see also related) from POTASSIUM. With more cross-referencing in the vocabulary, including references to and from main heading/subheading combinations, it would not be unreasonable to expect that a biomedical practitioner with a need for information could be led to the construction of reasonably sophisticated strategies. We cannot expect, of course, that he will achieve the same level of expertise that the trained analyst does.

Let us consider how an on-line system might assist a biomedical specialist in the construction of conventional search strategies, using terms with the Boolean AND, OR and NOT operators. Later we will consider alternatives to this conventional approach.

If the professional person is ever to approach the level of search sophistication achieved by the experienced analyst, he must be led by the hand. One possible approach is the use of spontaneously generated displays. When a user enters a descriptor he should have the option of seeing, immediately displayed, the term itself and its relations to other terms in the vocabulary. Ideally he should be given a display of the hierarchy (tree) in which the term appears, the terms it is used for and the "related terms" to which it is linked by see also references. He should also be given tallies to show term usage in indexing. A second display, on request, would yield the term definition, if one exists, and a list of permissible subheadings. The full capabilities of the computer can be exploited to generate additional useful displays that are not provided in MEDLARS at present. For example, one display would show, for any main heading entered, which other headings had been used most frequently with it in indexing, with statistics on the frequency of co-occurrence of the two terms. Another display would show how frequently each subheading had been used with a particular main heading.

The important thing is that at least some of these displays should be generated spontaneously unless previously suppressed. We are giving the user searching cues, helping him to decide whether or not his original terms are those best representing the subject of his search, and leading him to other related terms in as painless a way as possible. Spontaneous display of this type cannot be effected conveniently on a typewriter terminal: it requires the use of a video console. Probably the console should use a split-screen approach: one portion of the screen will contain the spontaneous displays, the other will be a storage area in which the searcher records selected terms. He should be able to incorporate into his strategy terms from any of the displays by the simple depression of a key or use of a light pen.

The spontaneously generated displays could avoid many of the missed approaches that occurred in the searches analyzed in this study. Loss of relevant documents in various searches occurred, for example, in the following cases:

(a) The term EMBOLISM was used but EMBOLISM, FAT was not.
(b) The term INFANT was used but INFANT, NEWBORN and INFANT, PREMATURE were not.
(c) The term ANEURYSM, MYCOTIC was used but the related term ENDOCARDITIS, BACTERIAL was not.
(d) The term HYPERTHYROIDISM was used but the more specific term GOITER, EXOPHTHALMIC was not.

These examples of missed search approaches could possibly have been avoided by various term displays--alphabetical, cross-referenced, and hierarchical. The hierarchical Tree Structures is potentially the most useful tool for the searcher, experienced or inexperienced. For the inexperienced searcher, however, it needs adequate cross-referencing between categories (e.g., lungs to lung diseases, respiratory tract physiology, and so on). Hedges or search fragments are also potentially very valuable aids to the inexperienced searcher and should also be capable of display under various conditions.

The storage area of the screen should be designed to resemble a simple search strategy form with the logical operators already displayed. For example, all terms in a line, by definition, are considered to be in an or relationship. Lines can be combined (anded) by a simple key depression. The system should be capable of accepting entry terms and should compensate automatically for most errors of spelling and punctuation. The check-off boxes now appearing on the MEDLARS request form could be reproduced on the screen to help the user usefully delimit the scope of his search. In the design of programs to guide the user in the construction of effective strategies, many useful procedures may be drawn from the highly related techniques of computer-assisted instruction. Some useful work in this area has already been conducted and reported by Caruso.[2]

Ultimately, however, on-line retrieval systems for use by people who are not information specialists should be designed to: (1) allow input of natural-language requests, and (2) avoid the necessity for inputting requests as formal search statements with Boolean operators. The advantages of the natural-language approach include the obvious one that the user does not need to learn the terms of some restricted vocabulary and the less obvious one that natural-language statements tend to represent true information needs better than statements that have been influenced by the logical and linguistic constraints of a system. This was revealed clearly in the results of a full MEDLARS evaluation.[1] Natural-language search statements can be used in systems employing controlled vocabularies for indexing purposes. However, an extensive entry vocabulary is needed to allow the necessary mapping to take place.

The use of Boolean algebra for querying computer-based retrieval systems may have been a mistake. The mistake arose through the way that early computer-based systems developed. That is, they were viewed as more mechanized versions of semi-manual systems such as those employing edge-notched cards or the optical coincidence principle. With these forerunner systems Boolean algebraic search expressions were necessary and they responded simply by retrieving or not retrieving references that matched the search strategy exactly. There was no possibility for "partial match" and the retrieval of lists of references ranked according to degree of match with the search strategy. Computer-based systems, on the other hand, have no reason to be so restricted. A computerized retrieval system can use algorithms that will rank documents according to degree of match with the search strategy. Boolean search equations are unnecessary and are probably undesirable in mechanized retrieval systems. Retrieval by term weighting, as described by Brandhorst[3] and by Sommar and Dennis,[4] among others, gives greater flexibility and allows the use of ranking procedures. The term weights may be assigned by the searcher or, under certain conditions, they may be supplied automatically, as indicated by Williams[5] and by Curtice and Jones.[6]

Ideally, then, the user at an on-line terminal should be able to enter natural-language search statements and have these statements converted automatically to <u>MeSH</u> terms or combinations. For instance, the natural-language statement "Neonatal Grave's Disease" should be automatically convertible to GOITER, EXOPHTHALMIC/CONGENITAL, where the entry vocabulary recognizes "Grave's Disease" as equivalent to GOITER, EXOPHTHALMIC and "neonatal" as equivalent to the subheading CONGENITAL.

Completely automatic translation of natural-language search statements to controlled vocabulary terms requires the development, and use within the computer, of "conceptual groups" of related terms. These conceptual groups may comprise terms hierarchically related and terms related by <u>see also</u> references, and they will include synonyms and other entry terms. For example, suppose a user enters the request "Freeze Storage of Skin and Kidneys." The word "freeze" is mapped automatically, via the entry vocabulary, to a conceptual group that includes the descriptors FREEZE DRYING, FREEZING and REFRIGERATION. "Storage" is mapped to a storage and preservation group that includes the descriptors PRESERVATION and TISSUE PRESERVATION. Likewise, "skin" and "kidneys" are mapped to appropriate conceptual groups. Once this mapping has taken place, the search algorithm looks for documents that have been indexed under various combinations of the terms occurring in the conceptual groups. The more of these terms a document profile contains the more relevant the document is presumed to be and the higher up it will appear in the printout. With search statements that factor into several conceptual groups, the selectin algorithm will create a requirement level that essentially establishes a retrieval threshold and cut-off point. For example, in the statement above there are four conceptual groups represented. The search algorithms will only allow retrieval of documents in which at least two of these conceptual groups are present. Documents in which three or four are present would obviously appear much higher on the ranked output.

Note that in this search mode the user can query the system in English sentence form without the need to use controlled terms and without the need to combine words in any Boolean expression. Curtice and Jones[6] have described an on-line, interactive system of this general type: users may enter natural-language search statements and these are subjected to an "automatic indexing routine." All necessity for specifying logical combinations of terms is removed from the requester. The search strategy is constructed automatically and the query terms are automatically weighted.

It is also desirable that an on-line retrieval system should incorporate certain unconventional searching methods. In particular, a user should be able to enter the system by citations to documents that he already knows to be relevant to his information need. In this case the search algorithm will retrieve the index term profiles of the known relevant items, will automatically develop a weighted search profile, and will examine the other profiles in the data base to find further documents indexed like those already known to be relevant. The same procedure can be used, obviously, once the on-line user has discovered one or more relevant items by more conventional search techniques.

In the above paragraphs I have attempted to indicate directions in which I feel on-line search procedures in MEDLARS, and indeed other information systems, should go in the long run. This is not to imply that the ELHILL system fails. I have said that it is a very satisfactory system of its own kind. However, we must not necessarily assume that future on-line systems, even those in immediate sight, should follow the same pattern. We should always look for ways of improving retrieval systems and making them more attractive to potential users. The philosophy that "the system is used, therefore it is good" is a very shallow one. We must not assume that a system having appeal today will always retain this appeal. There is a certain novelty factor about AIM-TWX that is at least partly responsible for the very favorable acceptance it has received in most quarters. But novelty wears off and system designers cannot afford to rest too long on their laurels. In the past users have been required to adapt to the information system. In the future systems must be designed that adapt to the user.

REFERENCES

1. U.S. National Library of Medicine. Evaluation of the MEDLARS Demand Search Service. Prepared by F. W. Lancaster. Washington, D.C., U.S. Dept. of Health, Education and Welfare, Public Health Service, 1968.

2. Caruso, D.E. "Tutorial Programs for Operation of On-Line Retrieval Systems," Journal of Chemical Documentation, 10:98-105, May 1970.

3. Brandhorst, W.T. "Simulation of Boolean Logic Constraints Through the Use of Term Weights," American Documentation, 17:145-46, July 1966.

4. Sommar, Helen G., and Dennis, Don E. "A New Method of Weighted Term Searching with a Highly Structured Thesaurus." In American Society for Information Science. Proceedings of the Annual Conference. Vol. 6. Westport, Conn., Greenwood Pub. Co., 1969, pp. 193-98.

5. Williams, John H., Jr. BROWSER--An Automatic Indexing On-Line Text Retrieval System (AD 693 143). Gaithersburg, Md., International Business Machines Corporation, 1969.

6. Curtice, R.M., and Jones, P.E. An Operational Interactive Retrieval System. Cambridge, Mass., Arthur D. Little, Inc., 1969.

VITA

F. Wilfrid Lancaster is currently Associate Professor of Library Science and Director of the Program in Biomedical Librarianship at the University of Illinois Graduate School of Library Science. He is a Fellow (by thesis) of the Library Association of Great Britain, and he attended the Newcastle upon Tyne School of Librarianship.

Before coming to the University of Illinois Mr. Lancaster was Senior Research Assistant, ASLIB Cranfield Project (1963); Head, Systems Evaluation Group, Herner & Co. (1964-65); Information Systems Specialist, National Library of Medicine (1965-68); and Director of Information Retrieval Services, Westat Research Inc. (1969).

Mr. Lancaster's fields of special interest are information storage and retrieval, medical libraries and industrial libraries. He is the author of <u>Information Retrieval Systems</u> (John Wiley, 1968), <u>Vocabulary Control for Information Retrieval</u> (Information Resources Press, 1972) and numerous reports and articles in the field of information sciences.

* *

Papers in this series are issued irregularly and no more often than monthly. Standing orders can be established for $5.00 which will cover ten papers to be sent as they are issued. Orders for any ten or more copies of back papers will be charged $5.00. Individual copies of back issues still in print are distributed free upon request. The Occasional Papers deal with any aspect of librarianship and consist of manuscripts which are too long or too detailed for publication in a library periodical or which are of specialized or temporary interest. Manuscripts for inclusion in this series are invited, and should be sent to: Occasional Papers, Publications Office--215 Armory Building, University of Illinois Graduate School of Library Science, Champaign, Illinois 61820.

* *